

細胞システムのモデリングとパラメータ推定に関する研究

中荃隆¹、仲隆²、木村周平³、畠山眞里子¹

¹ 理化学研究所 ゲノム科学総合研究センター

² 九州産業大学 情報科学部 知能情報学科

³ 鳥取大学 工学部 知能情報工学科

1. 緒言

細胞は細胞外の環境に応じて細胞の増殖・分化・細胞死を制御し、細胞数や大きさ、形質を調節している。このような細胞運命の決定は、細胞外から受けるシグナルとその情報を遺伝子へと伝搬する細胞内シグナル伝達系によって厳密に制御されている[1]。通常は細胞が属する個体の利益にかなうように制御されるはずのシグナル伝達のシステムに何らかの異常が生じると、癌などの疾病を引き起こす。細胞内シグナル伝達系の理解は単に生命現象の理解に留まらず、疾病予防や薬剤開発との関連も深い[2]。この系は、シグナル伝達因子として機能するタンパク質個々の生化学的な性質とそれら集団としての活性化の連鎖反応の時空間動態によって特徴付けられる。前者に対しては、タンパク質立体構造解析、*in vivo*、*in vitro*におけるタンパク質間相互作用解析などにより、知見が積み上げられている。一方で、後者に対しては、未だ決定的な生物学的、情報学的方法論は確立していない。特に、シグナル伝達系をネットワークとして見たとき、そのノード(タンパク質)とリンク(相互作用関係)の接続状態(パスウェイ)を同定する情報学的手法が求められている。

本研究では、細胞内シグナル伝達系のパスウェイ予測を数理モデルベースの計算論的アプローチで行うことを目的とする。今回用いたチャイニーズハムスター卵巣由来(CHO)細胞は細胞膜上に発現する受容体の種類を人工的に変えることで同一の細胞外刺激に対して、シグナルの量的に異なる応答を誘導することができる。この応答の違いに着目し、可能性として考えられるパスウェイ構造の中で、どのパスウェイ構造がこのシグナル量の違いを再現できるかを数理モデルのパラメータ同定問題を解くことで決定する手法を提案する。非線形微分方程式モデルのパラメータ同定は、非線形関数最適化問題に帰着されるため多くの計算量を必要とする。さらに、本研究では多数のモデル候補に対して最適化問題を解く必要があるため、PC クラスタを用いた並列計算を行うことで実時間で問題を解くことに成功した。

2. 問題設定

次の2種類の細胞株に対し、入力刺激として、上皮成長因子(EGF)を投与する。

- ・上皮成長因子(EGF)受容体のみを発現させた細胞(E11細胞)
- ・EGF受容体とErbB4受容体を共発現させた細胞(E14細胞)

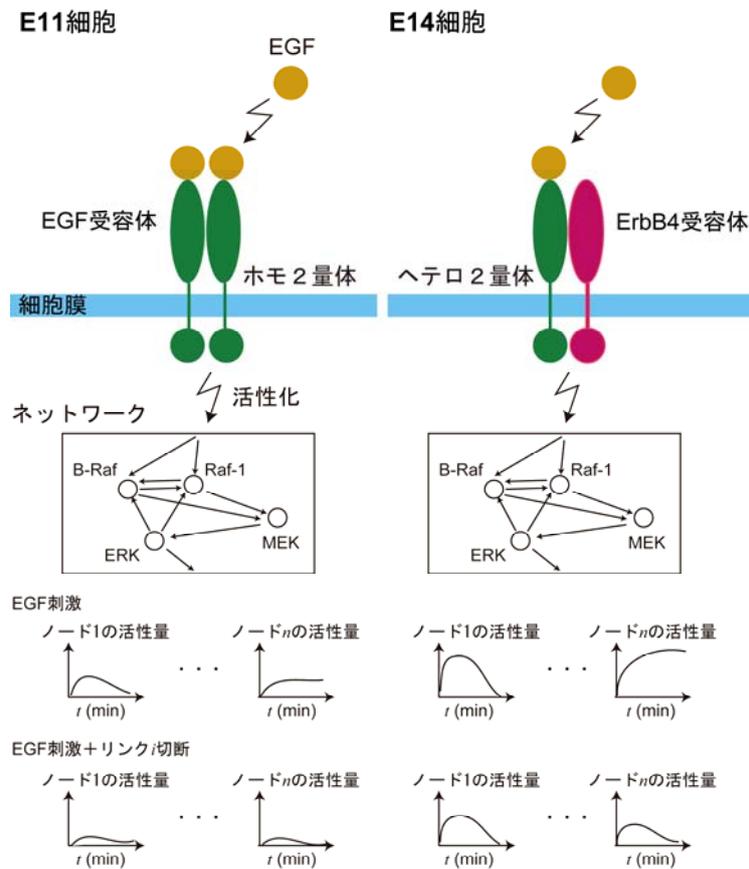


図1. EGF受容体誘導のシグナル伝達系

また、ネットワーク内の各タンパク質(ノード)の活性化応答を 30 分に渡って観測する(図 1)。本研究では、観測対象のタンパク質として Raf-1、B-Raf、MEK、ERK タンパク質を選んだ。これらは EGF 刺激によってシグナル伝達系内で応答する主要なタンパク質である。E11 細胞では EGF 受容体がホモ 2 量体を形成し、活性化するのに対し、E14 細胞ではホモ 2 量体に加えて、ヘテロ 2 量体も形成することが知られている(図1)。この膜表面上で生じる活性化メカニズムの違いは、下流のネットワークへ異なる入力シグナルとして伝搬し、各ノードの活性化量・パターンの違いとなって観測される。さらに、阻害剤を用いて特定のリンクを切断し、各ノードの活性化量・パターンがどのように変化するかも観測する。これらのデータを用いて次節で述べるようにパスウェイ構造の予測を行う。

3. 方法論

3.1 ネットワーク構造の候補の選別

Raf-1、B-Rafを入力ノード、ERK を出力ノードとする。受容体の活性化は、2 つの入力ノードを活性化し、出力ノードの活性化を引き起こす。本研究で行った生化学実験、及び、先行研究結果より以下の事実が分かっている。

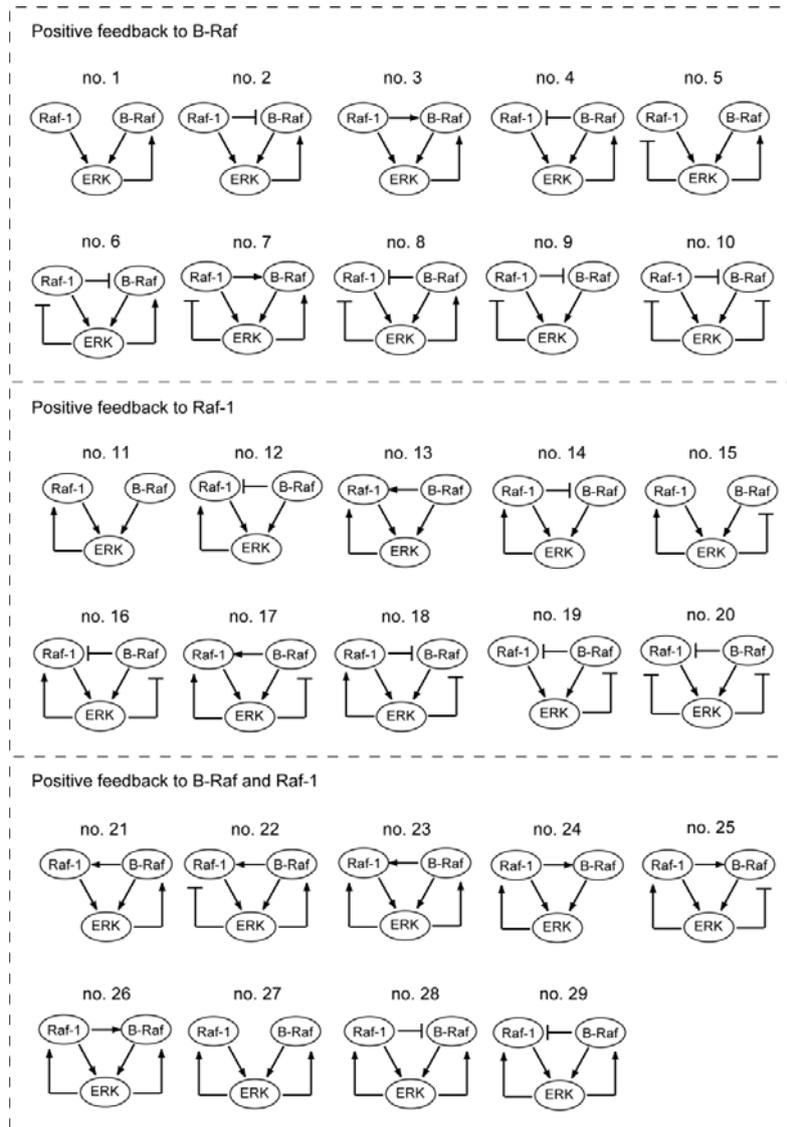


図 2. 29 種類のネットワーク構造

- 系内に ERK ノードから Raf-1/B-Raf ノードへの正のフィードバックが存在。
- MEK→ERK のリンクは既知。

このとき、以上の条件を満たすネットワーク構造は 29 種類となる(図 2*)。

3.2 ネットワーク構造の推定アルゴリズム

29 種類のネットワーク構造それぞれに対応する常微分方程式モデルを作成する。各モデルは 9 個の状態変数、リンクの本数に対応して 12~14 個のミカエリス・メンテン反応速度式とその反応速度を決める 24~28 個のキネティックパラメータで記述される。ネットワークのシステムとしての動態はパスウェイ構造とキネティックパラメータによって特徴付けられるため、もしある微分方程式モデルが実験的に観測される各ノードの活

* 図を見易くするため、MEKノードを省略して描かれている。

性状態を再現できるとすると、そのモデルに対応するネットワーク構造は実際の細胞内のネットワーク構造の特徴を捕えていると考えられる。従って、ネットワーク構造の推定問題は数理モデルのパラメータ推定問題へと帰着される。以下に、ネットワーク構造の推定の手順を示す。詳しくは、参考文献[3]を参照されたい。

1. モデル i ($i = 1, \dots, 29$) に対して、キネティックパラメータの推定を行う。実験的に測定した各ノードの活性化時系列パターンを教師信号とし、モデル内のノードの活性化時系列パターンとの誤差が最小になるようにパラメータを推定する。推定アルゴリズムとして、本研究では木村らによって提案された非線形関数最適化器 GLSDC を用いた[4]。
2. 手順 1 をパラメータの初期探索開始点を変えながら 3 回試行し、モデル i で推定誤差が最小となるパラメータを選ぶ。
3. モデル i に対して、判定条件として以下の事象を実験値とシミュレーション値で比較する。このとき、誤差が設定した閾値以内に収まっているかどうかをテストする。
 - (3-1) ノード j の活性化時系列パターン
 - (3-2) リンク k を切断した際のノード j の活性ピークの変化量

4. 結果

今回の実験で測定したノードは B-Raf、MEK、ERK ノードである。また、MEK-ERK 間のリンクを阻害する薬剤を用いたとき、MEK ノードの活性ピークがどのように変化するかを測定した。以上の測定は E11 細胞、及び、E14 細胞に対して行われた。ただし、B-Raf ノードの実測データは定量性が十分ではなかったため、E14 細胞を用いた実験の 30 分間に渡るパターン(持続的な/一過性の活性)のみを判定条件として用いた。手順 1~2 において、GLSDC による 29 個の微分方程式モデルのパラメータ推定に必要とされる計算時間は次式となる。

●1CPU (AMD Opteron 2.2GHz)

$$40 \text{ 時間} \times 29 \text{ モデル} \times 3 \text{ 試行} = \underline{3480 \text{ 時間}} (145 \text{ 日})$$

●160CPU (16CPU での並列計算を同時に 10 ジョブ行う) :

$$8 \text{ 時間} (16\text{CPU}) \times 29 \text{ モデル} \times 3 \text{ 試行} / 10 \text{ 並列ジョブ} = \underline{69.6 \text{ 時間}} (2.9 \text{ 日})$$

本研究では、PC クラスターを用いて 160CPU で並列計算を実行した。1 モデル 1 試行の推定にかかる時間は 16CPU 使用時におよそ 1/5 倍程度に留まったが、同時に 10 ジョブ並列実行することで、トータルでの計算時間は 1/50 倍に短縮することができた。手順 3 の実行結果の抜粋(モデル 1~7)を図 3 に示す。ここで、判定条件を満たした場合には“pass”を、満たさなかった場合には“fail”を各小グラフのタイトルに示した。全 29 モデルに対する結果より、モデル 4 とモデル 29 のみが全判定条件を“pass”することがわかった。図 2 に示されるように、モデル 4 とモデル 29 の違いは、ERK→Raf-1 へ

の正のフィードバックリンクの有無である。しかし、このリンクに対応する反応速度式のパラメータを見ると非常に小さな値が割り当てられており、実質的にモデル 29 はモデル 4 と等価であることが分かった。以上の結果より、推定対象のネットワークはモデル 4 のパスウェイ構造で特徴付けられることが分かった。

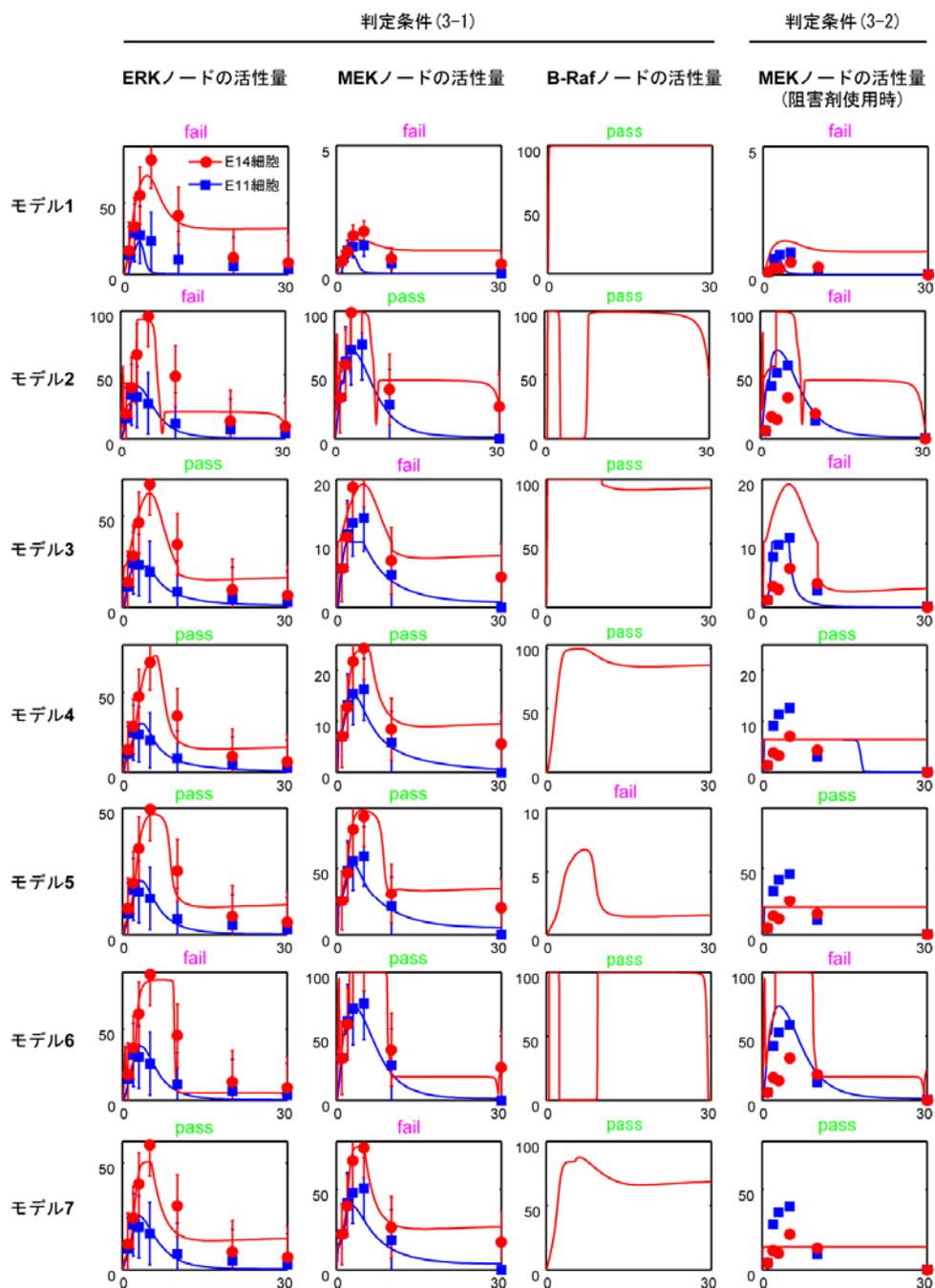


図3. 手順3の実行結果 (モデル1~7)

5. まとめ

細胞内シグナル伝達系の量的変化の特徴は、各タンパク質の立体構造やタンパク質間相互作用の結合特性だけで捉えることはできず、ネットワーク内の各タンパク質（ノード）の活性動態、つまり、フィードバックを伴ったネットワークのシステムとしての動態を考慮する必要がある。本研究では、CHO 細胞内で機能する比較的小規模なネットワークをターゲットとして、そのパスウェイ構造を推定した。その方法論の核はパスウェイ構造の推定問題を数理モデルのパラメータ推定問題に帰着させる点である。数理モデルはダイナミクスを表現する微分方程式で記述されており、その状態変数は各ノードの活性化量であるため、パスウェイ構造が推定されると、そのネットワークのシステムとしての動態をも推定したことになるため、さらなる解析へと研究を進めることが可能となる。参考文献[3]では、推定されたパスウェイ構造を持つシステムに対して、定常状態応答解析を行い、ERK ノード活性化に対してクリティカルなパスウェイを予測している。本研究で提案されたアルゴリズムは大規模なモデルへと拡張することが可能であるが、パラメータ推定問題の次元数も大きくなるため実時間で計算を終わらせるためには相応の並列計算機が必要となると考えられる。

謝辞

今回、実験系として用いた CHO 細胞は、理研 GSC タンパク質基盤研究グループ横山茂之プロジェクトディレクター、同研究グループタンパク質システム研究チーム白水美香子上級研究員よりご提供頂いた。この場を借りて感謝申し上げます。また、実験を行った理研 GSC システム情報生物学研究グループ情報伝達システムバイオロジー研究チーム湯本範子氏に対して、ここに記して謝意を表す。本研究では、RSCC システム (RIKEN Super Combined Cluster) の PC クラスタを利用して並列計算を行った。

参考文献

- [1] 中荃隆、仲隆、畠山眞里子、“数理モデルを用いた細胞内シグナル伝達系の解析～ポストゲノム時代におけるシステムバイオロジー研究の最前線”、計測と制御、vol.47, no.2, pp.132-138, 2008.
- [2] 畠山眞里子、中荃隆、仲隆、“RTK シグナル伝達系のシステムバイオロジー～数理解析の基礎と創薬への応用”、実験医学、vol.24, no.16, pp.2530-2535, 2006.
- [3] Nakakuki T., Yumoto N., Naka T., Shirouzu M., Yokoyama S. & Hatakeyama M., “Topological Analysis of MAPK Cascade for Kinetic ErbB Signaling”, *PLoS ONE*, 2008 (in press).
- [4] Kimura S. & Konagaya A., “High Dimensional Function Optimization using a new Genetic Local Search suitable for Parallel Computers”, *Proc. of the 2003 Int. Conf. on Systems, Man, and Cybernetics*, 335-342, 2003.